

DOCUMENT NO. 156-57  
RESEARCH LABORATORY OF ELECTRONICS  
MASSACHUSETTS INSTITUTE OF TECHNOLOGY  
CAMBRIDGE 39, MASSACHUSETTS, U.S.A.

*Loan Copy  
only*

#5

# PERIODIC SAMPLING OF STATIONARY TIME SERIES

JOHN P. COSTAS

TECHNICAL REPORT NO. 156

MAY 16, 1950

RESEARCH LABORATORY OF ELECTRONICS  
MASSACHUSETTS INSTITUTE OF TECHNOLOGY

The research reported in this document was made possible through support extended the Massachusetts Institute of Technology, Research Laboratory of Electronics, jointly by the Army Signal Corps, the Navy Department (Office of Naval Research) and the Air Force (Air Materiel Command), under Signal Corps Contract No. W36-039-sc-32037, Project No. 102B; Department of the Army Project No. 3-99-10-022.

MASSACHUSETTS INSTITUTE OF TECHNOLOGY  
Research Laboratory of Electronics

Technical Report No. 156

May 16, 1950

PERIODIC SAMPLING OF STATIONARY TIME SERIES

John P. Costas

Abstract

When measuring the mean or average value of a random time function by periodic sampling some error is to be expected since only a finite number of samples are taken. An expression for the mean-square error of a sample mean is derived. The variation of error with changes in total sampling time and spacing between samples is discussed. An expression for the error involved in determining the mean by continuous integration is derived and the concept of sampling efficiency is introduced. A brief discussion of the probability of error is presented.

---

\_\_\_\_\_

## PERIODIC SAMPLING OF STATIONARY TIME SERIES

### 1. Introduction

The continued study of the statistical theory of communication as originated by Wiener and developed by Lee has led the communication engineer to a closer study of the statistical nature of certain stationary random processes. One of the simplest, yet one of the most important, properties of a random time function is its average or mean value. For example, correlation functions which are of great importance in the new theory, involve the average of a product of two time functions. It is to be expected that the design of electronic correlators (1) would involve sampling theory. Many other examples may be given to demonstrate the importance of the theory of sampling to current research.

It is the purpose of this paper to examine the error involved in determining the mean of a random function by sampling when restriction is placed upon either the number of samples to be taken or the total time available for sampling. In addition, only periodic sampling will be considered here\*, i.e. the samples are equally spaced from one another in time. Although similar developments may be found in the literature\*\* it is believed that most engineers engaged in statistical studies will find some of these results new and, it is hoped, useful.

### 2. Variance of the Sample Mean

Consider the stationary random process represented by  $Z(t)$  of Fig. 1. Let the total time available for sampling be  $T$  and the time interval between samples be  $t_0$ . The individual sample values are designated by  $Z_j$ , where  $j = 1, 2, 3, \dots, n$ . The sample mean will then be given by\*\*\*

$$\bar{Z} \triangleq \frac{1}{n} (Z_1 + Z_2 + \dots + Z_n) \quad (1)$$

where  $t_0 n = T$  as shown in Fig. 1.

---

\* In a practical sense this is hardly a restriction as almost all electronic sampling is periodic.

\*\* Equation 14 may be found in many places; for example in Wilks (3). A search of statistical literature has not yielded anything resembling Eq. 11 although Uspensky (4), p. 173, discusses a similar problem.

\*\*\*The symbol  $\triangleq$  reads, "equal by definition".

Now in general  $\bar{Z}$  will not be exactly equal to the mean of  $Z(t)$  since only a finite number of samples are taken over a finite period of time. Certainly if  $T$  and  $n$  are large, one expects  $\bar{Z}$  to be very nearly equal to the mean. Rather than examine one specific  $\bar{Z}$  it is useful to consider the

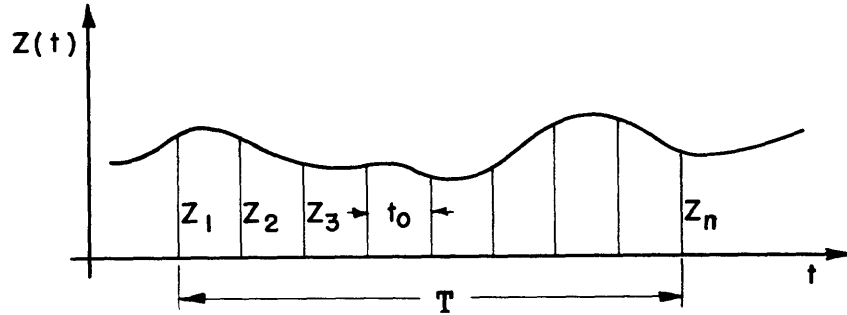


Figure 1.

results of many independent sample means of the type 1. Then we may consider the average squared error of a number of sample means as an indication of the sampling efficiency of a given  $T$  and  $n$ .

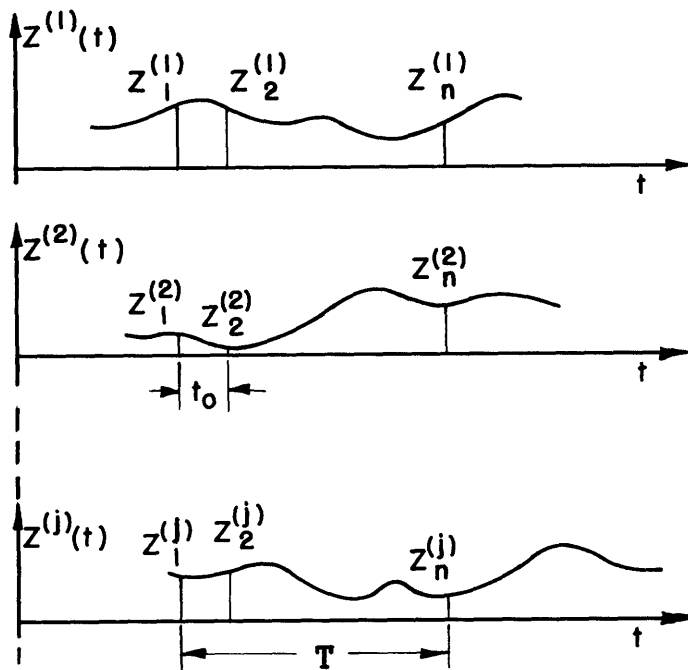


Figure 2.

It is therefore useful at this point to consider an ensemble of time functions  $Z(t)$  and a corresponding ensemble of sample means  $\bar{Z}$ , as shown in Fig. 2. The various members of the ensemble are indicated by the superscript. Each  $Z(t)$  is produced by independent but similarly prepared systems. This enables us to consider  $\bar{Z}$  and the individual  $Z_j$ 's as random variables.

Let the variable  $Z(t)$  have a variance  $\sigma^2$  and a mean value  $m$ . That is,

$$\left. \begin{aligned} \text{and} \quad E[Z(t)] &= m \\ E\left[\left\{Z(t) - m\right\}^2\right] &= E[Z^2(t)] - m^2 = \sigma^2 \end{aligned} \right\} \quad (2)$$

Let us now find the variance of  $\dot{Z}$ , which is designated by  $\sigma_{\dot{Z}}^2$ . This variance is of particular importance since it is in effect a measure<sup>2</sup> of the mean-square error of a set of independent sample means. This can be seen from Eq. 2a.

$$\sigma_{\dot{Z}}^2 = E[(\dot{Z} - m)^2] = E[\dot{Z}^2] - E^2[\dot{Z}] \quad (2a)$$

Since the mean of a sum is equal to the sum of the means, we may write

$$E[\dot{Z}] = \frac{1}{n} \cdot n E(Z) = m, \quad (3)$$

since

$$E[Z_1] = E[Z_2] = \dots = E[Z_n] = E[Z(t)] = m. \quad (4)$$

To find  $\sigma_{\dot{Z}}^2$  we now need only  $E[\dot{Z}^2]$ . From Eq. 1 it follows that

$$E[\dot{Z}^2] = \frac{1}{n^2} E\left[\sum_{r=1}^n Z_r \sum_{s=1}^n Z_s\right]$$

or

$$E[\dot{Z}^2] = \frac{1}{n^2} E\left[Z_1^2 + Z_2^2 + \dots + Z_n^2 + \sum_{r,s=1}^n{}' Z_r Z_s\right] \quad (5)$$

where the prime indicates that all terms  $Z_r Z_s$  having  $r = s$  are ignored. Considering each  $Z_1, Z_2, \dots, Z_n$  as a random variable we may write

$$E[Z_1^2] = E[Z_2^2] = \dots = E[Z_n^2] = E[Z^2(t)]. \quad (6)$$

Thus with the aid of Eq. 2 we may rewrite Eq. 5 as

$$E[\dot{Z}^2] = \frac{1}{n} (\sigma^2 + m^2) + \frac{1}{n^2} E\left[\sum_{r,s=1}^n{}' Z_r Z_s\right]. \quad (7)$$

Before the second term on the right-hand side of Eq. 7 may be evaluated it is necessary to define  $\phi(\tau)$ , the autocorrelation function of  $Z(t)$ .

$$\phi(\tau) \triangleq \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T Z(t) Z(t + \tau) dt = E[Z(t) Z(t + \tau)] \quad (8)$$

Since Eq. 7 involves terms of the kind  $E[Z_r Z_{r+k}]$ , we may use Eq. 8 and write

$$E[Z_1 Z_{1+k}] = \phi(kt_0) \quad (9)$$

If

$$\sum_{r,s=1}^n{}' Z_r Z_s$$

is expanded one finds exactly  $2(n - k)$  terms  $Z_r Z_{r+k}$ , where  $k = 1, 2, \dots, n - 1$ . Thus Eq. 7 may be written as

$$E[\dot{Z}^2] = \frac{(\sigma^2 + m^2)}{n} + \frac{1}{n^2} \sum_{k=1}^{n-1} 2(n - k) \phi(kt_0). \quad (10)$$

The variance of  $\dot{Z}$ ,  $\sigma_{\dot{Z}}^2 = E[Z^2] - m^2$  may now be written as

$$\sigma_{\dot{Z}}^2 = \frac{\sigma^2}{n} + \frac{1}{n^2} \sum_{k=1}^{n-1} 2(n - k) \phi(kt_0) + \frac{m^2(1 - n)}{n}. \quad (11)$$

Thus Eq. 11 represents the mean-square error of a sample mean composed  $n$  samples spaced  $t_0$  seconds apart. If the total sampling time  $T$  is specified, then

$$nt_0 = T. \quad (11a)$$

An interesting special case of Eq. 11 occurs when  $t_0$  is sufficiently large so that each sample is statistically independent of all other samples. In such a case  $\phi(kt_0)$  is equal to  $m^2$  for all  $k$ . (This may be verified by reference to Eq. 8.) Equation 11 then becomes

$$\sigma_{\dot{Z}}^2 = \frac{\sigma^2}{n} + \frac{2m^2}{n^2} \sum_{k=1}^{n-1} (n - k) + \frac{m^2(1 - n)}{n}. \quad (12)$$

Since

$$\sum_{k=1}^{n-1} (n - k) = 1 + 2 + 3 + \dots + n - 1 = \frac{n(n - 1)}{2} \quad (13)$$

Eq. 12 becomes

$$\sigma_{\dot{Z}}^2 = \frac{\sigma^2}{n} \quad (14)$$

which is the well-known expression for the variance of a sample mean composed of independent samples.

### 3. Minimization of the Variance $\sigma_{\dot{Z}}^2$

Equation 11 must now be examined in order to determine the effect upon  $\sigma_{\dot{Z}}^2$  of variations in  $n$  and  $T$ . In this section it will be convenient to designate the variance of the sample mean by  ${}_n\sigma_{\dot{Z}}^2$  where  $n$  indicates the size of the sample mean.

Consider the difference  ${}_{n+1}\sigma_{\dot{Z}}^2 - {}_n\sigma_{\dot{Z}}^2$  as obtained from Eq. 11.



$${}_{n+1}\sigma_z^2 - {}_n\sigma_z^2 = -\frac{(\sigma^2 + m^2)}{n(n+1)} + 2 \sum_{k=1}^n \phi\left(\frac{kT}{n}\right) \left[ \frac{(2n+1)k - n(n+1)}{n^2(n+1)^2} \right] \quad (15)$$

Inspection of Eq. 15 shows that  ${}_{n+1}\sigma_z^2 - {}_n\sigma_z^2$  approaches zero as  $n$  approaches infinity. This indicates that Eq. 11 approaches a minimum value as  $n$  increases,  $T$  being kept constant. Before continuing further we define a new correlation function  $\phi'(\tau)$ .

$$\phi'(\tau) \triangleq \phi(\tau) - m^2 \quad (16)$$

Therefore  $\phi'(\tau) \rightarrow 0$  as  $\tau \rightarrow \infty$ . Equation 11 may now be rewritten as

$${}_n\sigma_z^2 = \frac{\sigma^2}{n} + \frac{2}{n^2} \sum_{k=1}^{n-1} (n-k) \phi'(kt_0) \quad (17)$$

or

$${}_n\sigma_z^2 = \frac{\sigma^2}{n} + 2 \sum_{kt_0=t_0}^{T-t_0} \frac{t_0}{T} \left(1 - \frac{kt_0}{T}\right) \phi'(kt_0) . \quad (17a)$$

As  $n \rightarrow \infty$  we have:  $t_0 \rightarrow \Delta\tau$ ,  $kt_0 \rightarrow \tau$ . Letting  ${}_{\infty}\sigma_z^2 = \lim_{\tau \rightarrow \infty} {}_n\sigma_z^2$ , we may write with the aid of Fig. 3:

$${}_{\infty}\sigma_z^2 = \frac{2}{T} \int_{\tau=0}^{\tau=T} \left(1 - \frac{\tau}{T}\right) \phi'(\tau) d\tau . \quad (18)$$

Equation 18 represents the limiting value of the variance of the sample mean as the number of samples in a fixed period  $T$  is increased beyond limit.

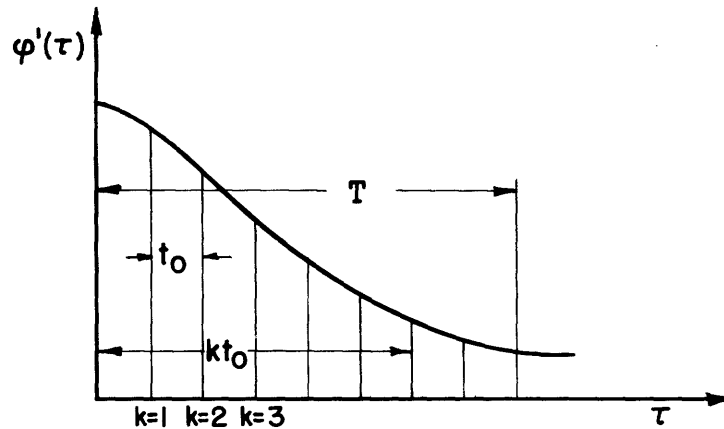


Figure 3.

We may now define a sampling efficiency  $\eta$ , where

$$\eta \triangleq \frac{\sigma_{\dot{Z}}^2}{n \sigma_{\dot{Z}}^2} = \frac{\frac{2}{T} \int_0^T (1 - \frac{\tau}{T}) \phi'(\tau) d\tau}{\frac{\sigma^2}{n} + \frac{2}{n^2} \sum_{k=1}^{n-1} (n-k) \phi'(kt_0)} . \quad (19)$$

Thus in a given situation a computation of  $\eta$  would be of aid in choosing  $n$ . For example, if  $\eta$  were found to be 0.95 for  $n = 1000$  there would be little point in designing equipment to sample at  $n = 2000$ .

From the properties of correlation functions (2) one can show that  $\sigma_{\dot{Z}}^2 \rightarrow 0$  as  $T \rightarrow \infty$ .

#### 4. Probability of Error

Consider a random variable  $x$ , with variance  $\sigma_x^2$  and mean  $\bar{x}$ . The Tchebycheff inequality states that (5)

$$P(|x - \bar{x}| > a) \leq \frac{\sigma_x^2}{a^2} \quad (20)$$

which says that the probability of  $x$  differing from its mean by a magnitude greater than  $a$ , is less than or equal to  $\sigma_x^2/a^2$ . Letting  $\dot{Z}$  be  $x$  we may write at once

$$P(|\dot{Z} - m| > a) \leq \frac{\sigma_{\dot{Z}}^2}{a^2} . \quad (21)$$

Thus the probability that  $\dot{Z}$  differs from the true mean by an amount greater than  $a$  is at most equal to  $\sigma_{\dot{Z}}^2/a^2$ .

## References

1. Y. W. Lee: Detection of Periodic Signals in Random Noise by the Method of Correlation, Technical Report No. 157, Research Laboratory of Electronics, M.I.T., to be published.
2. A. Khintchine: Korrelationstheorie der Stationären Stochastischen Prozesse, Math. Ann. 109, 604 (1934).
3. S. S. Wilks: Mathematical Statistics (Princeton Univ. Press, 1943).
4. J. V. Uspensky: Introduction of Mathematical Probability (McGraw-Hill, N.Y., 1937).
5. H. Cramer: Mathematical Methods of Statistics (Princeton Univ. Press, 1946).

